

# A LSE and Sparse Message Passing-Based Channel Estimation for mmWave MIMO Systems

Chongwen Huang\*, Lei Liu<sup>†</sup>, Chau Yuen\*, Sumei Sun<sup>‡</sup>

\*Singapore University of Technology and Design, 487372, Singapore

<sup>†</sup>State Key Lab of ISN, Xidian University, Xi'an, 710071, China

<sup>‡</sup>Institute for Infocomm Research (I2R), A\*STAR, 138632, Singapore

E-mail: chongwen\_huang@mymail.sutd.edu.sg

**Abstract**—In this paper, we propose a novel channel estimation algorithm based on the Least Square Estimation (LSE) and Sparse Message Passing algorithm (SMP), which is of special interest for Millimeter Wave (mmWave) systems, since this algorithm can leverage the inherent sparseness of the mmWave channel. Our proposed algorithm will iteratively detect exact the location and the value of non-zero entries of sparse channel vector without its prior knowledge of distribution. The SMP is used to detect exact the location of non-zero entries of the channel vector, while the LSE is used for estimating its value at each iteration. Then, the analysis of the Cramer-Rao Lower Bound (CRLB) of our proposed algorithm is given. Numerical experiments show that our proposed algorithm has much better performance than the existing sparse estimators (e.g. LASSO), especially when mmWave systems have massive antennas at both the transmitters and receivers. In addition, we also find that our proposed algorithm converges to the CRLB of the genie-aided estimation of sparse channels in just a few turbo iterations.

## I. INTRODUCTION

Millimeter wave (mmWave) is receiving tremendous interest by academia, industry, and government for future 5G cellular systems [1], [2]. The main reason is that the majority of our current wireless communications systems operating in the microwave spectrum (i.e., <6 GHz) are by now crowded and limited bandwidth. MmWave can take full advantage of spectrum from 30 GHz to 300 GHz and provide beyond 2 GHz of bandwidth [3]. However, the larger communication spectrum means more challenges. One of the main challenges, mmWave signal propagation is impaired by severe signal attenuation. Recent urban model experiments show that path losses are 40 dB worse at 28 GHz compared to 2.8 GHz [3].

One way to overcome this severe signal attenuation of mmWave propagation paths is to improve beamforming gain by increasing number of transmit antennas and receive antennas. This leads to an important feature of mmWave channels, which is very sparse in both the angle and time domains [2], [4]. Recently research measurements [3] show that mmWave channels typically exhibit only 3-4 scattering clusters in dense-urban Non-Line-Of-Sight and Line-Of-Sight (LOS) environments. However, conventional MIMO channel estimation methods cannot be applied directly in mmWave systems since these methods did not account for mmWave channel sparsity. For example, in [5], [6], two joint channel estimation schemes for massive MIMO systems were proposed, and they were both based on message-passing iterative algorithms, which can

reduce the complexity. However, they did not specialize for mmWave systems. This prompts the need to design efficient channel estimation techniques for the mmWave.

For the sparse channel estimation, a well-known approach is the LASSO [7] with a tuning parameter that trades between the sparsity and measurement-fidelity of the solution. In [8], the authors investigated three algorithms for sparse channel estimation, and they were Approximate Maximum Likelihood Estimator (AMLD-SE), Iterative Detection/Estimation With Threshold (ITD-SE) and the Sphere Detection (SD-SE). In [9], an alternating minimization method for sparse channel estimation was proposed. Simulation results showed that ITD-SE had a better performance than others. However, it needs an adaptive threshold selection strategy to keep the performance, which reduces its flexibility. Recently, approaches to deal with sparse channels have been proposed in literatures by resorting to Compressed Sensing (CS) theory. In [10], an adaptive CS algorithm for mmWave channel estimation was proposed. This algorithm is very suitable for the mmWave, while it relies on the design of a multi-resolution beamforming codebook.

In this paper, we propose a novel channel estimation scheme based on the Least Square Estimation (LSE) and Sparse Message Passing (SMP) algorithm. Compared with previously proposed sparse channel estimation, ours can yield a better performance since it not only can take full advantage of the inherent sparseness of mmWave channel, but also can leverage the virtues of LSE and SMP algorithm. Firstly, we introduce an intermediate virtual channel representation that is attractive since it can capture the essence of physical modeling and provides simple geometric interpretation of the scattering environment([see [4], [11]]). Then, our proposed algorithm involves to iteratively detect exact the location and the value of non-zero entries. The SMP is used to detect exact the location of non-zero entries of the channel vector, while the LSE is used for estimating its value at each iteration. Furthermore, we give the analysis of the Cramer-Rao Lower Bound (CRLB) of our proposed algorithm. Numerical simulations show that our algorithm exhibits far better performance than the conventional LSE estimator as well as existing sparse estimator (e.g. LASSO). In addition, we also find that this algorithm needs only four turbo iterations to nearly achieve the CRLB.

**Notation:**  $a$  is a scalar,  $\mathbf{a}$  is a vector and  $\mathbf{A}$  is a matrix.  $\mathbf{A}^T$ ,  $\mathbf{A}^H$ ,  $\mathbf{A}^{-1}$ ,  $\mathbf{A}^\dagger$  and  $\|\mathbf{A}\|_F$  represent transpose, Hermitian

(conjugate transpose), inverse, pseudo-inverse and Frobenius norm of a matrix  $\mathbf{A}$ , respectively.  $\mathbf{A} \otimes \mathbf{B}$  denotes the Kronecker product of  $\mathbf{A}$  and  $\mathbf{B}$ , and  $\text{vec}(\mathbf{A})$  is a vector stacking all the columns of  $\mathbf{A}$ .  $\text{diag}(\mathbf{a})$  is a diagonal matrix with the entries of  $\mathbf{a}$  on its diagonal, and  $\text{diag}(\mathbf{A})$  is a vector that its entries come from the diagonal elements of  $\mathbf{A}$ .

## II. SYSTEM MODEL

We consider a mmWave communication system that has  $N_t$  transmit antennas,  $N_r$  receive antennas, and the narrow band baseband received signal can be written as follows:

$$\mathbf{y} = \mathbf{H}\mathbf{s} + \mathbf{n} \quad (1)$$

where  $\mathbf{H} \in \mathbb{C}^{N_r \times N_t}$  is the channel matrix,  $\mathbf{s} \in \mathbb{C}^{N_t \times 1}$  is the transmitted signal,  $\mathbf{y} \in \mathbb{C}^{N_r \times 1}$  is the received signal, and  $\mathbf{n} \sim \mathcal{N}(0, \sigma^2 \mathbf{I})$  is the Gaussian noise corrupting the received signal. Since mmWave channels are expected to have limited scattering, we adopt a geometric channel model with  $L$  scatterers. Each scatterer is further assumed to contribute a single propagation path between transmitters and receivers [3]. Under this model, the channel  $\mathbf{H}$  can be expressed as:

$$\mathbf{H} = \sum_{l=1}^L \alpha_l \mathbf{a}_r(\theta_l) \mathbf{a}_t^H(\phi_l) \quad (2)$$

where  $\alpha_l$  is the gain of the  $l$ th path,  $\phi_l \in [0, 2\pi]$ , and  $\theta_l \in [0, 2\pi]$  denote the  $l$ th path's azimuth angles of departure and arrival of transmitters and receivers respectively. Finally,  $\mathbf{a}_r(\theta_l)$  and  $\mathbf{a}_t(\phi_l)$  are the antenna array response vectors at receivers and transmitters respectively. If A Uniform Planar Array (UPA) is used,  $\mathbf{a}_t(\phi_l)$  can be written as:

$$\mathbf{a}_t(\phi_l) = \frac{1}{\sqrt{N_t}} [1, e^{j\frac{2\pi}{\lambda} \sin(\phi_l)}, \dots, e^{j(N_t-1)\frac{2\pi}{\lambda} \sin(\phi_l)}]^T \quad (3)$$

where  $\lambda$  is the signal wavelength, and  $d$  is the distance between antenna elements. The array response vectors at the receivers,  $\mathbf{a}_r(\theta_l)$ , can be written in a similar fashion. Then, the channel can be written in a more compact form as:

$$\mathbf{H} = \mathbf{A}_r \text{diag}(\boldsymbol{\alpha}) \mathbf{A}_t^H \quad (4)$$

where  $\boldsymbol{\alpha} = \sqrt{\frac{N_r N_t}{\rho}} [\alpha_1, \alpha_2, \dots, \alpha_L]^T$ . The matrices

$$\mathbf{A}_t = [\mathbf{a}_t(\phi_1), \mathbf{a}_t(\phi_2), \dots, \mathbf{a}_t(\phi_L)] \quad (5)$$

and

$$\mathbf{A}_r = [\mathbf{a}_r(\theta_1), \mathbf{a}_r(\theta_2), \dots, \mathbf{a}_r(\theta_L)] \quad (6)$$

contain the transmitter and receiver array response vectors. Finally,  $\rho$  is the path-loss between the transmitters and receivers.

The mmWave channel is usually dominated by the LOS, or consists of a single reflected path [12]. Large antenna arrays are deployed to get beamforming gain in order to combat the high path loss. Hence, we usually have  $L \ll \min\{N_r, N_t\}$ .

Then, we adopt an intermediate virtual channel representation that keeps the essence of physical modeling without its complexity, provides a tractable linear channel characterization, and offers a simple and transparent interpretation on the effects of scattering and array characteristics on channel capacity. The finite dimensionality of the signal space allows the virtual channel model to be expressed as:

$$\mathbf{H}_v = \mathbf{W}_r^H \mathbf{H} \mathbf{W}_t \quad (7)$$

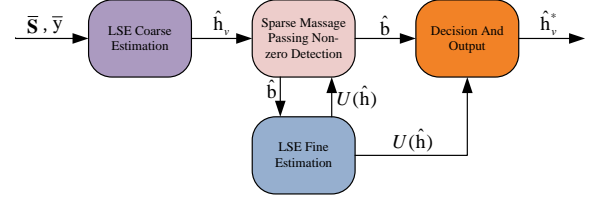


Fig. 1. The processing for proposed LSE-SMP algorithm, which consists of three phases: Coarse Estimation, Sparse Message Passing Detection, and Fine Estimation.

fig:env

Note that  $\mathbf{W}_r \in \mathbb{C}^{N_r \times N_r}$  and  $\mathbf{W}_t \in \mathbb{C}^{N_t \times N_t}$  are channel-invariant unitary Discrete Fourier Transform (DFT) matrices [13]. We recast (1) by the virtual channel representation as:

$$\mathbf{y} = \mathbf{W}_r \underbrace{\mathbf{W}_r^H \mathbf{H} \mathbf{W}_t}_{:=\mathbf{H}_v} \mathbf{W}_t^H \mathbf{s} + \mathbf{n} \quad (8)$$

$$= \mathbf{W}_r \mathbf{H}_v \mathbf{W}_t^H \mathbf{s} + \mathbf{n}$$

The fourier transformation  $\mathbf{W}_t$  and  $\mathbf{W}_r$  can be seen as a mapping from the antenna domain onto an angular domain and the entries of the matrix  $\mathbf{H}_v$  can be interpreted as the channel gains between  $N_t$  transmit and the  $N_r$  receive beams. In this paper, we assume that the array response vectors are along the directions defined in the DFT matrices. This means that  $\mathbf{H}_v$  is perfectly sparse, i.e., has exactly  $L$  non-zero entries.

Assuming the channel is time-invariant in the blocks  $t \in \{1, \dots, T\}$ . Then,  $\mathbf{Y} \triangleq [\mathbf{y}_1, \dots, \mathbf{y}_T]$ ,  $\mathbf{S} \triangleq [\mathbf{s}_1, \dots, \mathbf{s}_T]$ , and  $\mathbf{N} \triangleq [\mathbf{n}_1, \dots, \mathbf{n}_T]$ . We rewritten the channel model as:

$$\mathbf{Y} = \mathbf{W}_r \mathbf{H}_v \mathbf{W}_t^H \mathbf{S} + \mathbf{N} \quad (9)$$

where  $\mathbf{Y} \in \mathbb{C}^{N_r \times T}$ ,  $\mathbf{S} \in \mathbb{C}^{N_t \times T}$ , and  $\mathbf{N} \in \mathbb{C}^{N_r \times T}$ . We define  $\mathbf{X} \triangleq \mathbf{W}_t^H \mathbf{S}$ . Then, we recast the (9) as:

$$\mathbf{Y} = \mathbf{W}_r \mathbf{H}_v \mathbf{X} + \mathbf{N} \quad (10)$$

Then, we can factorize the (10) [10] as:

$$\text{vec}(\mathbf{Y}) = \text{vec}(\mathbf{W}_r \mathbf{H}_v \mathbf{X} + \mathbf{N}) \quad (11)$$

$$\stackrel{(a)}{=} (\mathbf{X}^T \otimes \mathbf{W}_r) \text{vec}(\mathbf{H}_v) + \text{vec}(\mathbf{N})$$

where (a) follows from the equality  $\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A}) \text{vec}(\mathbf{B})$ . By defining the  $\bar{\mathbf{S}} \triangleq \mathbf{X}^T \otimes \mathbf{W}_r$ , we simply the (11) in a compact fashion.

$$\bar{\mathbf{y}} = \bar{\mathbf{S}} \mathbf{h}_v + \bar{\mathbf{n}} \quad (12)$$

where  $\bar{\mathbf{y}} \in \mathbb{C}^{N_r T \times 1}$ ,  $\bar{\mathbf{S}} \in \mathbb{C}^{N_r T \times N_t N_r}$ ,  $\mathbf{h}_v \in \mathbb{C}^{N_t N_r \times 1}$  and  $\bar{\mathbf{n}} \in \mathbb{C}^{N_r T \times 1}$ . The mmWave channel estimation problem is simplified to estimate virtual channel vector  $\mathbf{h}_v$  by giving the equivalent training matrix  $\bar{\mathbf{S}}$  and the observed signal  $\bar{\mathbf{y}}$ .

## III. SPARSE CHANNEL ESTIMATION

In this section, we present a novel channel estimation algorithm based on the LSE and SMP algorithm, which is named LSE-SMP, and it consists of three phases: Coarse Estimation, Message Passing Detection, and Fine Estimation. This algorithm is shown in the Fig. 1. Since there is no priori knowledge of  $\mathbf{h}_v$ , we initially adopt LSE method to obtain its coarse estimation. Then, we consider the estimation of non-zero positions in the channel vector  $\mathbf{h}_v$  as a detection problem, and propose a SMP algorithm to find these non-zero positions. Again, we apply LSE method by leveraging the estimated non-zero positions to obtain the fine estimation. The second step and third step will repeat until we obtain the steady estimation.

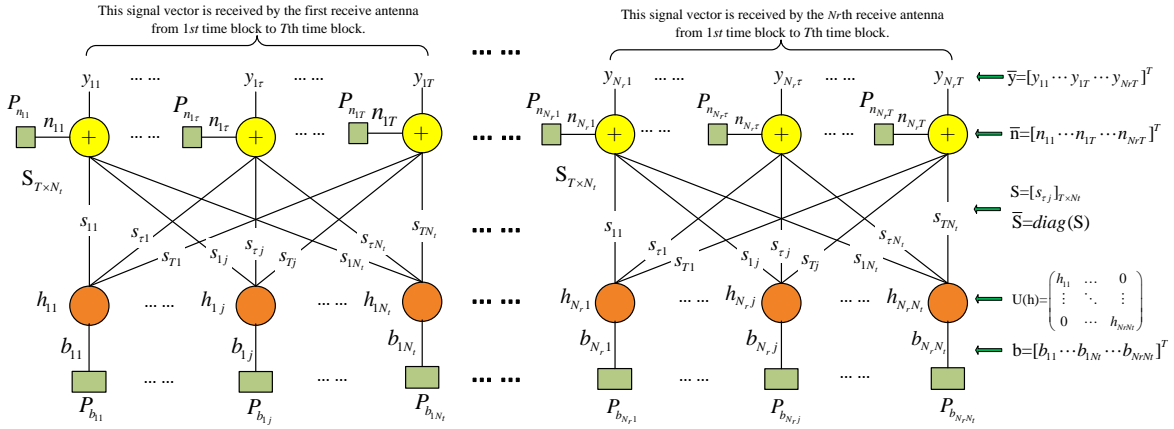


Fig. 2. Factor Graph Representation for Proposed Sparse Message Passing (SMP) Detection Algorithm

### A. LSE Coarse Estimation

To find Coarse Estimation  $\hat{\mathbf{h}}_v$  based on the observed vector  $\bar{\mathbf{y}}$ , with a Mean Square Error (MSE), we can solve the following Least Square (LS) problem:

$$\begin{aligned} \hat{\mathbf{h}}_v &= \arg \min_{\mathbf{h}_v} \|\bar{\mathbf{y}} - \bar{\mathbf{S}}\mathbf{h}_v\|_2^2 \\ &= [\bar{\mathbf{S}}^T \bar{\mathbf{S}}]^{-1} \bar{\mathbf{S}}^T \bar{\mathbf{y}} \end{aligned} \quad (13)$$

It is noted that it is optimal estimator in the sense of MSE when the estimator does not have any prior knowledge about neither the sparsity structure of  $\mathbf{h}_v$  (i.e. the location of non-zero entries), nor its degree of sparsity (i.e.  $L$ ).

### B. Sparse Message Passing Algorithm

After we get the Coarse Estimation of  $\mathbf{h}_v$ , we propose a fast iterative algorithm to find the positions of non-zero entries. This algorithm is named Sparse Message Passing since it can take full advantage of channel sparsity and message passing.

1) *Factor Graph Representation of mmWave Channel*: In order to get better understanding of our proposed algorithm, we show factor graph representation of channel vector  $\mathbf{h}_v$  in the following. Firstly, we decompose the  $\mathbf{h}_v$  into a diagonal coefficient matrix  $\mathbf{U}(h_{ij}) (i \in [0, \dots, N_r], j \in [0, \dots, N_t])$  and a column array  $\mathbf{b}$ . The column array  $\mathbf{b} = [b_{ij}]_{N_r N_t \times 1}$  is called position vector, and it represents the positions of non-zero in virtual channel vector  $\mathbf{h}_v$ . The  $b_{ij} \in \{1, 0\}$  can be seen as a Bernoulli distribution. Then, the  $\mathbf{h}_v$  can be recast as:

$$\mathbf{h}_v = \underbrace{\begin{bmatrix} h_{11} & & & 0 \\ & \ddots & & \\ & & h_{1N_t} & \\ 0 & & & h_{N_r N_t} \end{bmatrix}}_{=\mathbf{U}(\mathbf{h})} \underbrace{\begin{bmatrix} b_{11} \\ \vdots \\ b_{1N_t} \\ \vdots \\ b_{N_r N_t} \end{bmatrix}}_{=\mathbf{b}} \quad (14)$$

The equivalent training matrix  $\bar{\mathbf{S}}$  can be expressed as the following block diagonal matrix by designing  $\mathbf{X}$  and  $\mathbf{W}_r$ .

$$\bar{\mathbf{S}} = \text{diag}(\mathbf{S}) \quad (15)$$

where  $\bar{\mathbf{S}} \in \mathbb{C}^{N_r T \times N_r N_t}$ , and  $\mathbf{S}$  can be seen as a matrix of transmitted training sequences, which will be received by  $i$ th ( $i \in [1, \dots, N_r]$ ) receive antenna from 1 to  $T$ th time block. It can be written as:

$$\mathbf{S} = [s_{\tau j}]_{T \times N_t} \quad (16)$$

where  $s_{\tau j}$  represents the transmitted training sequences by  $j$ th ( $j \in [1, \dots, N_t]$ ) transmit antenna in the  $\tau$ th ( $\tau \in [1, \dots, T]$ ) time block. Then, we rewrite (12) as:

$$\begin{aligned} \underbrace{\begin{bmatrix} y_{11} & \dots & y_{1T} & \dots & y_{N_r T} \end{bmatrix}^T}_{\bar{\mathbf{y}}} &= \bar{\mathbf{S}}\mathbf{h}_v + \bar{\mathbf{n}} \\ &= \bar{\mathbf{S}}\mathbf{U}(\mathbf{h})\mathbf{b} + \bar{\mathbf{n}} \end{aligned} \quad (17)$$

According to the factor graph analysis rule, we plot a factor graph to represent above equations, and it is shown in the Fig. 2. The nodes  $(n_{11}, \dots, n_{N_r T})$  and  $(h_{11}, \dots, h_{N_r N_t})$  are named the sum and variable nodes respectively.

Our proposed SMP algorithm is considered for estimating positions of non-zero. It is similar to the belief propagation decoding process of low density parity check code, in which the output message called extrinsic information on each edge is calculated by the messages on the other edges that are connected with the same node [14].

2) *Message Update at Sum Nodes of SMP*: To analysis a sum node that is shown in the Fig. 3, we can know the received signal at the  $\tau$ th time block by the  $i$ th receive antenna, and it can be expressed as:

$$y_{i\tau} = \sum_{j=1}^{N_t} s_{\tau j} h_{ij} b_{ij} + n_{i\tau} \quad (18)$$

From assumption of channel model in the above section, we know that there are  $L$  non-zero in the vector of  $\mathbf{b}$ . Assuming that  $b_{ij}$  is independent and identically distributed (i.i.d.). Then, we can know the initial probability of the Bernoulli distribution, which can be denoted as:

$$\begin{cases} p_0(b_{ij} = 1) = \frac{L}{N_r N_t}, \\ p_0(b_{ij} = 0) = 1 - \frac{L}{N_r N_t}. \end{cases} \quad (19)$$

According to the law of large numbers, when  $N_t$  goes very large, we can know the term  $\sum_{j=1}^{N_t} s_{\tau j} h_{ij} b_{ij}$  can be approximated as Gaussian distribution [15]. When we compute the probability of  $p(b_{ij} = 1)$  from  $i\tau$  sum node to  $ij$  variable node, we consider the messages from other variable nodes to sum node  $i\tau$  as noise. This can be expressed as:

$$y_{i\tau} = \underbrace{s_{\tau j} h_{ij} p(b_{ij} = 1)}_{\text{desired}} + \underbrace{\sum_{m \neq j}^{N_t} s_{\tau m} h_{im} p(b_{im} = 1) + n_{i\tau}}_{\text{interference}} \quad (20)$$

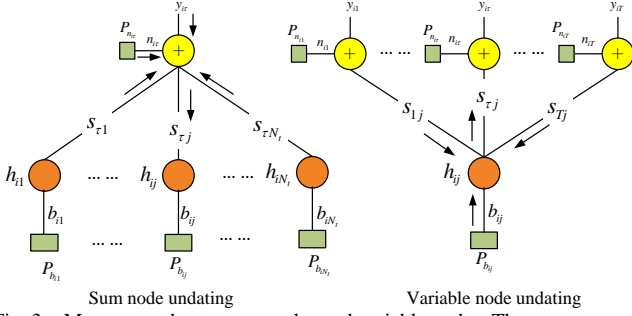


Fig. 3. Message update at sum nodes and variable nodes. The output message called extrinsic information on each edge is calculated by the messages on the other edges that are connected with the same node. For the Gaussian-Bernoulli sparse signals, the messages passing on each edge are the mean and variance of a Gaussian distribution, and the non-zero probability of a Bernoulli distribution.

Furthermore, we can compute its mean value and variance. The message update at the sum nodes is given by:

$$\begin{cases} e_{i\tau \rightarrow ij}^s(k) = \sum_{m \neq j} s_{\tau m} \hat{h}_{im} p_{im \rightarrow i\tau}^v(k-1), \\ v_{i\tau \rightarrow ij}^s(k) = \sum_{m \neq j} s_{\tau m}^2 p_{im \rightarrow i\tau}^v(k-1) \\ \quad \times (v_{h_{im}} + \hat{h}_{im}^2 (1 - p_{im \rightarrow i\tau}^v(k-1))) + \sigma_n^2. \end{cases} \quad (21)$$

where  $j, m \in \{1, 2, \dots, N_t\}$ ,  $\sigma_n^2$  is the variance of the Gaussian noise. In addition,  $p_{im \rightarrow i\tau}^v(k-1)$  denotes the probability message of passing from the  $im$  variable node to  $i\tau$  sum node at  $(k-1)$ th iteration [16]. Similarly,  $e_{i\tau \rightarrow ij}^s(k)$  and  $v_{i\tau \rightarrow ij}^s(k)$  denote the mean and variance of  $s_{\tau j} h_{ij}$  passing from the  $i\tau$  sum node to  $ij$  variable node at  $k$ th iteration respectively.  $\hat{h}_{im}$  and  $v_{h_{im}}$  denote the mean and variance of  $h_{im}$  that are estimated in the fine phase.

Once we get the mean and variance from the  $i\tau$  sum node to  $ij$  variable node, we can compute its statistical probability according to the  $s_{\tau j} h_{ij} p(b_{ij} = 1) \sim \mathcal{N}(e_{i\tau \rightarrow ij}^s(k), v_{i\tau \rightarrow ij}^s(k))$ . Firstly, we define the Gaussian probability density function as:

$$f(x|\mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (22)$$

Then, we can give the probability from the  $i\tau$  sum node to  $ij$  variable node as follow:

$$p_{i\tau \rightarrow ij}^s(k) = \frac{1}{1 + \frac{f(y_{i\tau} | e_{i\tau \rightarrow ij}^s(k), v_{i\tau \rightarrow ij}^s(k))}{f(y_{i\tau} | e_{i\tau \rightarrow ij}^s(k) + s_{\tau j} \hat{h}_{ij}, v_{i\tau \rightarrow ij}^s(k) + s_{\tau j}^2 v_{h_{ij}})}} \quad (23)$$

3) *Message Update at Variable Nodes of SMP*: In terms of the message update at variable nodes, we consider variable nodes as a broadcast process [14] and the message update at these nodes is given by:

$$p_{ij \rightarrow i\tau}^v(k) = \frac{1}{1 + \frac{\prod_{t \neq \tau} (1 - p_{it \rightarrow ij}^s(k-1)) \cdot p_0(b_{ij}=0)}{\prod_{t \neq \tau} p_{it \rightarrow ij}^s(k-1) \cdot p_0(b_{ij}=1)}} \quad (24)$$

where  $t, \tau \in \{1, 2, \dots, T\}$ , and,  $p_{it \rightarrow ij}^s(k-1)$  denotes the probability message passing from the  $it$  sum node to  $ij$  variable node at  $(k-1)$ th iterations.

### C. LSE Fine Estimation

Once the positions of the non-zero have been detected, the next step is to estimate the value of coefficient matrix  $\mathbf{U}(h)$ . For the problem, we propose a novel strategy based on LSE method. This strategy is to swap the position of  $h_{ij}$  and  $b_{ij}$

in the (17), so that we can get an accurate estimation by leveraging sparsity of  $\mathbf{b}$ . Rewriting the (17) as:

$$\bar{\mathbf{y}} = \bar{\mathbf{S}}\mathbf{U}(\mathbf{b})\mathbf{h} + \bar{\mathbf{n}} \quad (25)$$

where  $\mathbf{h} \in \mathbb{C}^{N_t N_r \times 1}$ . Similar with Coarse Estimation, the estimation of  $\mathbf{h}$  can be found by solving the following LS problem:

$$\hat{\mathbf{h}} = \arg \min_{\mathbf{h}} \{\|\bar{\mathbf{y}} - \bar{\mathbf{S}}\mathbf{U}(\hat{\mathbf{b}})\hat{\mathbf{h}}\|_2^2\} \quad (26)$$

where  $\hat{\mathbf{b}}$  is the vector obtained by SMP algorithm. Calculating the gradient of the above expression with respect to  $\hat{\mathbf{h}}$  and setting it to zero, we get the following estimator for  $\hat{\mathbf{h}}$ :

$$\begin{cases} \hat{\mathbf{h}}(k) = \hat{\mathbf{Q}}^\dagger(k-1)(\bar{\mathbf{S}}\mathbf{U}(\hat{\mathbf{b}}(k-1)))^T \bar{\mathbf{y}} \\ \mathbf{V}_h(k) = \sigma_n^2 (\bar{\mathbf{S}}\mathbf{U}(\hat{\mathbf{b}}(k-1)))^T \bar{\mathbf{S}}\mathbf{U}(\hat{\mathbf{b}}(k-1)) \end{cases} \quad (27)$$

where  $\hat{\mathbf{Q}}(k-1) = \mathbf{U}(\hat{\mathbf{b}}(k-1))\bar{\mathbf{S}}^T \bar{\mathbf{S}}\mathbf{U}(\hat{\mathbf{b}}(k-1))$ ,  $\mathbf{V}_h = [v_{h_{im}}]_{N_r N_t \times 1}$ ,  $i \in \{1, 2, \dots, N_r\}$  and  $m \in \{1, 2, \dots, N_t\}$ .  $\hat{\mathbf{h}}(k)$  and  $\mathbf{V}_h$  denotes the estimated value and variance of  $\mathbf{h}$  at  $k$ th iteration. After we obtain  $\hat{\mathbf{h}}(k)$ , this value will replace the  $\hat{\mathbf{h}}(k-1)$  for calculating  $\hat{\mathbf{b}}(k)$  by applying SMP algorithm.

### D. Decision and Output of LSE-SMP

When the MSE of the LSE-SMP meets the requirement or the number of iterations reaches the limit, we output:

$$\hat{b}_{ij} = \frac{1}{1 + \frac{\prod_{t=1}^T (1 - p_{it \rightarrow ij}^s(k)) \cdot p_0(b_{ij}=0)}{\prod_{t=1}^T p_{it \rightarrow ij}^s(k) \cdot p_0(b_{ij}=1)}} \quad (28)$$

and

$$\begin{cases} \hat{\mathbf{h}} = \hat{\mathbf{Q}}^\dagger(\bar{\mathbf{S}}\mathbf{U}(\hat{\mathbf{b}}))^T \bar{\mathbf{y}}, \\ \hat{\mathbf{h}}_v^* = \mathbf{U}(\hat{\mathbf{b}})\hat{\mathbf{h}}. \end{cases} \quad (29)$$

where  $\hat{\mathbf{h}}_v^*$  is the finally outputted channel estimation vector. It should be pointed out that the decision is made based on full information coming from all the sum nodes.

## IV. ANALYSIS OF CRAMER RAO BOUND

In this section, we give the analysis of Cramer-Rao bound and show that our proposed LSE-SMP algorithm is unbiased. We begin with conventional LSE that is applied for a deterministic and non-sparse channel vector  $\mathbf{h}_v$ . From [17] and recalling the signal model in (13), the CRLB is yielded as:

$$\text{CRLB}_{\text{LSE}} \geq \sigma_n^2 (\bar{\mathbf{S}}^T \bar{\mathbf{S}})^{-1} \quad (30)$$

Note that LSE is the Minimum Variance Unbiased Estimator (MVUE), if  $\mathbf{h}_v$  is deterministic and non-sparse.

Compared with the non-sparse case, the sparse situation is slightly more complicated. As we are interested in the lower bound for the estimation accuracy, we assume perfect knowledge of the non-zero positions, i.e.  $\mathbf{U}(\hat{\mathbf{b}}) = \mathbf{b}$ . Then, the next step is to verify the estimator is unbiased. Recalling the signal model in (25) and from the definition of unbiased estimator, we get:

$$\begin{aligned} E(\hat{\mathbf{h}}_v) &= E(((\bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^T \bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^\dagger (\bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^T \bar{\mathbf{y}})) \\ &= E(((\bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^T \bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^\dagger (\bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^T (\bar{\mathbf{S}}\mathbf{U}(\mathbf{b})\mathbf{h} + \bar{\mathbf{n}})) \\ &= \mathbf{U}(\mathbf{b})\mathbf{h} = \mathbf{h}_v \end{aligned} \quad (31)$$

So, it is a unbiased estimator. The next step is to compute its CRLB. As previously mentioned, the channel  $\mathbf{h}_v$  is a deterministic vector, we can get:

$$\bar{\mathbf{y}} \sim p(\bar{\mathbf{y}}; \mathbf{h}_v) = \mathcal{N}(\bar{\mathbf{S}}\mathbf{h}_v, \sigma_n^2 \mathbf{I}) \quad (32)$$

Then, we can compute the  $\frac{\partial \ln p(\bar{\mathbf{y}}; \mathbf{h}_v)}{\partial \mathbf{h}_v}$ ,  

$$\frac{\partial \ln p(\bar{\mathbf{y}}; \mathbf{h}_v)}{\partial \mathbf{h}_v} = \frac{1}{\sigma_n^2} [(\bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^T \bar{\mathbf{y}} - (\bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^T \bar{\mathbf{S}}\mathbf{U}(\mathbf{b})\mathbf{h}_v] \quad (33)$$

Then, we obtain the following expression for the Fisher Information Matrix (FIM):

$$\mathbf{I}(\mathbf{h}_v) = -E\left(\frac{\partial^2 \ln p(\bar{\mathbf{y}}; \mathbf{h}_v)}{\partial \mathbf{h}_v^2}\right) = \frac{1}{\sigma_n^2} (\bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^T \bar{\mathbf{S}}\mathbf{U}(\mathbf{b}) \quad (34)$$

We note that  $\bar{\mathbf{S}}\mathbf{U}(\mathbf{b})$  has the rank no larger than  $L$  due to the multiplication of  $\bar{\mathbf{S}}$  by  $\mathbf{U}(\mathbf{b})$ . The matrices  $\bar{\mathbf{S}}\mathbf{U}(\mathbf{b})$  and  $(\bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^T$  have some all zero columns (and rows), so it is singular. For this type singular matrix, it need to meet the following constraint [18], otherwise our proposed estimator (29) has infinite variance that renders the CRLB useless. Before we analyze this constraint, we firstly compute the following key identity:

$$\begin{aligned} \mathbf{G} &= ((\bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^T \bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^\dagger (\bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^T \bar{\mathbf{S}}\mathbf{U}(\mathbf{b}) \\ &= \text{diag}(\mathbf{b}) \neq \mathbf{I}_{N_r N_t} \end{aligned} \quad (35)$$

This constraint is given by:

$$\mathbf{G} = \mathbf{G}\mathbf{I}(\mathbf{h}_v)\mathbf{I}(\mathbf{h}_v)^\dagger \quad (36)$$

plugging the (34) and (35) into (36), we obtain  $\mathbf{G} = \mathbf{G}\mathbf{I}(\mathbf{h}_v)\mathbf{I}(\mathbf{h}_v)^\dagger = \mathbf{U}(\mathbf{b})\mathbf{U}(\mathbf{b})^T$  that holds. This means that the variance of our proposed estimator is finite. Since the FIM  $\mathbf{I}(\mathbf{h}_v)$  in (34) is singular, the expression for the CRLB can be computed following [18] which yields:

$$\begin{aligned} \text{CRLB}_{\text{LSE-SMP}} &\geq \mathbf{G}\mathbf{I}(\mathbf{h}_v)^\dagger \mathbf{G}^T \\ &= \sigma_n^2 (\bar{\mathbf{S}}\mathbf{U}(\mathbf{b}))^T \bar{\mathbf{S}}\mathbf{U}(\mathbf{b}) \end{aligned} \quad (37)$$

## V. NUMERICAL RESULTS

In this section, we investigate the performance of the our proposed algorithm using Monte-Carlo simulations, comparing with the LASSO sparse estimator and CRLB. We demonstrate that ours proposed LSE-SMP algorithm provides significant performance gains over existing techniques. In particular, we show that our proposed algorithm specialize in the mmWave channel estimation. Furthermore, we conduct numerical studies to investigate the impact of channel sparse ratios and iterations.

### A. Setup

For our numerical study, we considered the channel estimation problem in a  $32 \times 64$  mmWave MIMO system. The value and positions of non-zero elements in original channel vector  $\mathbf{h}_v$  were both generated by random way with a distribution following  $\mathcal{N}(0, \sqrt{10})$ . Throughout, we considered  $\text{SNR} \triangleq E\{\|\bar{\mathbf{S}}\|_F^2 / \|\bar{\mathbf{n}}\|_F^2\}$  in the interval  $[-10, 40\text{dB}]$ ; sparsity ratio was calculated by  $\eta = L/(N_r \times N_t)$ ; and the performance metric was the Normalized Mean Square Error (NMSE), given by  $E\{\frac{\|\hat{\mathbf{h}}_v - \mathbf{h}_v\|_F^2}{\|\mathbf{h}_v\|_F^2}\}$ .

### B. Performance Comparison

Fig. 4 shows the average channel-estimation NMSE of our proposed LSE-SMP algorithm and LASSO under the sparse ratio at 0.7%. Additionally, we also compute the CRLB for the classical LSE and our proposed LSE-SMP estimator. Our proposed LSE-SMP estimator obtains the better NMSE performance than that of LASSO and LSE. As expected, the

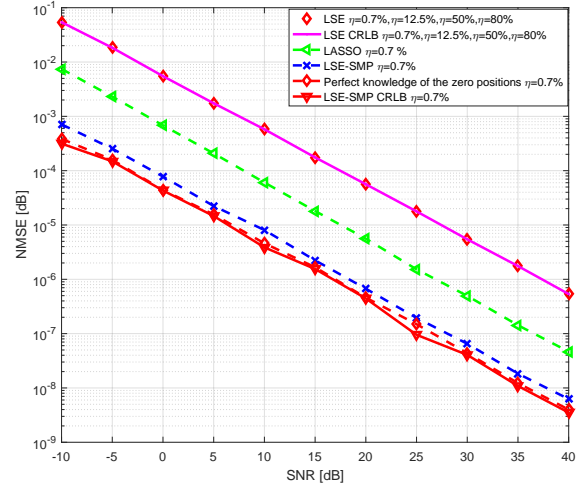


Fig. 4. Average NMSE performance comparison of proposed LSE-SMP, conventional LSE, and LASSO channel estimates versus SNR for the sparsity ratio at  $\eta = 0.7\%$  with the 5 iterations. It also shows CRLB of LSE and proposed LSE-SMP.

CRLB for our proposed LSE-SMP is the lowest, and this result is also consistent with that of classical LSE estimator with perfect knowledge of the non-zero positions. It is also seen from the Fig.4 that the gap between LSE-SMP and LSE-SMP CRLB is much smaller, about  $1.5\text{dB}$ . This is partly due to the errors in the detection of the non-zero positions in sparse message passing phase and partly to the fact that all our detection strategies rely on a coarse (and noisy) initial estimate of the channel.

### C. Effect of Sparsity Ratio

For further investigating the effect of sparsity ratio to our proposed algorithm, we fixed the iteration times  $= 5$ , and changed the sparsity ratio  $\eta$  from 0.7% to 80%. Note that different sparsity ratios  $\eta$  will leads to different LSE-SMP CRLBs. Simulations show that the NMSE performance of the LSE-SMP is very close to its CRLB in the each of sparsity ratios. In the Fig. 5, we just give the LSE-SMP CRLB at the sparse ratio  $\eta = 0.7\%$ . We can see that the NMSE performance of LSE is consistent LSE CRLB, and it is also the upper bound of LSE-SMP. These verify the analysis of LSE-SMP in the IV section. It means that the NMSE performance of LSE-SMP will be better with decreasing of sparse ratios as shown in the Fig. 5, essentially because LSE-SMP is able to exploit the sparsity of the channel. To be specific, The NMSE performance of LSE keep unchange under different sparsity ratios. On the other hand, the NMSE performance of LSE-SMP will increase with the decreasing of sparsity ratios. This means that LSE-SMP scheme will perform better especially when channel is very sparse.

### D. Effect of Iterations

Fig. 6 shows the average channel-estimation NMSE performance for LSE-SMP algorithm under several turbo iterations with sparsity ratio  $\eta = 3.1\%$ . The result shows that after the second turbo iteration, the NMSE performance of LSE-SMP performs a significant improvement. Additionally, we also find that the gap of the NMSE performance between the adjacent



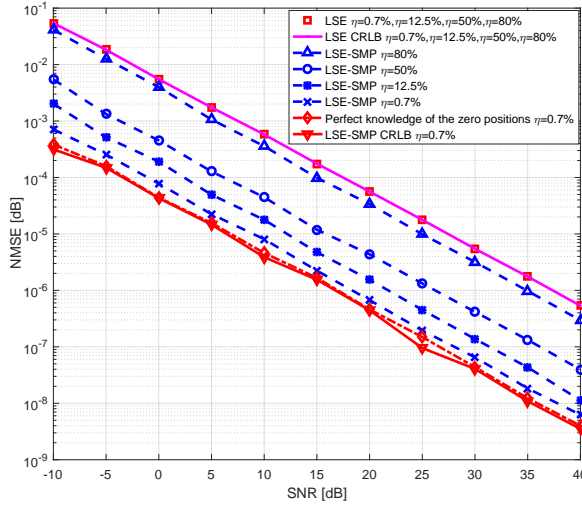


Fig. 5. Average NMSE performance comparison of LSE-SMP, LSE and their CRLB channel estimates versus SNR for different sparsity ratios  $\eta \in \{0.7\%, 12.5\%, 50\%, 80\%\}$  with the iteration = 5. It can be seen that LSE CRLB is the upper bound and LSE-SMP CRLB is the low bound of our proposed LSE-SMP.

iterations for LSE-SMP algorithm will be decreasing with increasing of iterations. After fourth turbo iteration, the NMSE performance have no significant improvement and it is very close to our analyzed LSE-SMP CRLB. This demonstrates that the convergence speed of the LSE-SMP algorithm is fast.

## VI. CONCLUSION

In this paper, a novel channel estimation algorithm for mmWave MIMO systems was proposed, which can take full advantage of the inherent sparseness of mmWave channel. This algorithm leverages the virtues of the SMP and LSE algorithm. The CRLB for the proposed LSE-SMP algorithm was analyzed. Simulation experiments verified that our proposed algorithm had much better performance than the existing sparse estimators, especially when channel is very sparse. Furthermore, it was also shown that the proposed algorithm needs only several turbo iterations to achieve CRLB. In spite of this, it still has the high complexity due to the computation of inverse matrixes in coarse and fine estimation phases. In the following, we plan to develop an approach to replace the LSE coarse and fine estimation without loss of global performance of the algorithm. Another limitation in our paper is the assumption that virtual channel vector  $\mathbf{h}_v$  has exactly  $L$  non-zero entries. In the future work, we will also consider to relax the assumption.

## REFERENCES

- [1] T. Rappaport, S. Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. Wong, J. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work" *IEEE Access*, vol. 1, pp. 335C349, 2013.
- [2] R. W. Heath Jr., N. G. Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An Overview of Signal Processing Techniques for Millimeter Wave MIMO Systems," to appear in *IEEE Journal of Selected Topics in Signal Processing*, Apr. 2016.
- [3] T. S. Rappaport, G. R. MacCartney, Jr., M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design" *IEEE Transactions on Communications*, vol. 63, no. 9, pp. 3029-3056, Sept. 2015.

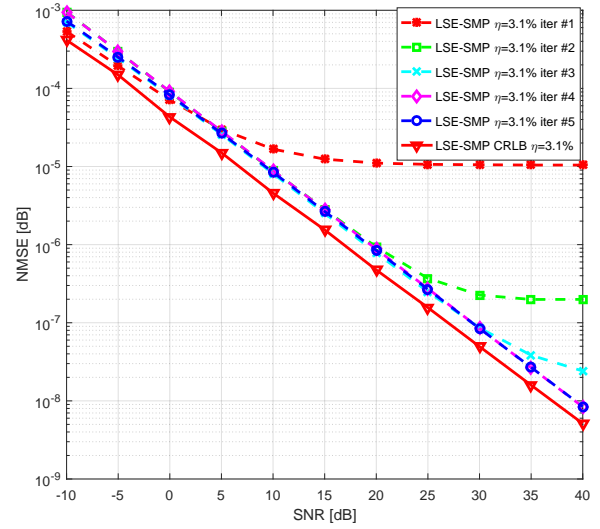


Fig. 6. Average NMSE performance comparison of LSE-SMP and its CRLB channel estimates versus SNR for different turbo iterations with the sparsity ratio  $\eta = 3.1\%$ . After the fourth iteration, the NMSE performance of LSE-SMP is converging to its CRLB.

- [4] P. Schniter and A. M. Sayeed, "A Sparseness-Preserving Virtual MIMO Channel Model," in *Proc. Conf. on Information Sciences and Systems*, (Princeton, NJ), pp. 36-41, Mar. 2004.
- [5] Y. Zhu, D. Guo and M. L. Honig, "A message-passing approach for joint channel estimation, interference mitigation, and decoding," *IEEE Transactions on Wireless Communications*, vol. 8, no. 12, pp. 6008-6018, Dec. 2009.
- [6] S. Wu, L. L. Kuang, Z. Y. Ni, D. Huang, Q. H. Guo and J. H. Lu, "Message-Passing Receiver for Joint Channel Estimation and Decoding in 3D Massive MIMO-OFDM Systems" submitted to *IEEE Transactions on Wireless Communications*, Jan. 2016.
- [7] R. Tibshirani, "Regression shrinkage and selection via the lasso" *J. Roy. Statist. Soc. B*, vol. 58, no. 1, pp. 267-288, 1996.
- [8] C. Carbonelli, S. Vedantam and U. Mitra, "Sparse Channel Estimation with Zero Tap Detection," *IEEE Transactions on Wireless Communications*, vol. 6, no. 5, pp. 1743-1763, May 2007.
- [9] R. Niazadeh, M. Babaie-Zadeh, and C. Jutten, "An alternating minimization method for sparse channel estimation," in *Proc. 9th Int. Conf. Latent Variable Anal. Signal Separation*, 2010, pp. 319C327.
- [10] A. Alkhateeb, O. El Ayach, G. Leus and R. W. Heath, "Channel Estimation and Hybrid Precoding for Millimeter Wave Cellular Systems," in *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 831-846, Oct. 2014.
- [11] A. M. Sayeed, "Deconstructing multi-antenna fading channels," *IEEE Trans. on Signal Processing*, pp. 2563C2579, Oct. 2002.
- [12] J. Mo, P. Schniter, N. Gonzalez-Prelcic, and R. W. Heath, Jr., "Channel estimation in millimeter wave MIMO systems with one-bit quantization" in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Nov. 2014.
- [13] P. Schniter and A. Sayeed, "Channel Estimation and Precoder Design for Millimeter-Wave Communications: The Sparse Way," in *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Nov. 2014.
- [14] L. Liu, Y. Li, Y. Su, and Y. Sun, "Quantize-and-Forward Strategy for Interleave-Division Multiple-Access Relay Channel," *IEEE Transactions on Vehicular Technology* 65(3): 1808-1814, 2016.
- [15] L. Liu, C. Yuen, Y. L. Guan, Y. Li and Y. P. Su, "Convergence Analysis and Assurance Gaussian Message Passing Iterative Detection for Massive MU-MIMO Systems," *IEEE Transactions on Wireless Communications*, to be published.
- [16] L. Liu, C. Yuen, Y. L. Guan, Y. Li and C. W. Huang, "Gaussian Message Passing Iterative Detection for MIMO-NOMA Systems With Massive Users," *IEEE Global Communications Conference (GlobeCom2016)*, Washington, DC USA, 2016.
- [17] A. Van den Bos, *Parameter estimation for scientists and engineers*. John Wiley and Sons, 2007.
- [18] P. Stoica and T. L. Marzetta, "Parameter estimation problems with singular information matrices," *IEEE Trans. Signal Processing*, vol. 49, no. 1, pp. 87-89, Jan. 2001.